Secure Sequence-based Localization for Wireless Networks

Bhaskar Krishnamachari and Kiran Yedavalli

Department of Electrical Engineering-Systems
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089
{bkrishna, kyedaval}@usc.edu

Summary. We present a unique sequence-based approach to localization that allows for automatic error correction. In dense deployments, this technique can provide accurate localization even in the face of malicious falsification of reference signals.

1 Introduction

Providing mobile devices with accurate location information is a fundamental service in many current and envisioned deployments of wireless networks, including wireless local area networks, ad hoc networks and wireless sensor networks. Besides being useful in itself in providing end users with location awareness, localization can be a key building-block for other wireless network protocols such as those for routing, sleep-scheduling, call-admission, etc. As we enter a world where wireless devices are deployed pervasively in industrial and military settings as well as public and private buildings, there is increasing interest in providing robust localization techniques that are not only functional (in terms of providing the desired level of accuracy and speed), but also secure with respect to possible malicious attacks.

In the abstract, of course, there are a large number of different security threats and concerns in a wireless network. In some settings, of greatest concern is protecting the privacy and confidentiality of localization, ensuring that the location of unknown nodes in the network is not revealed to outsiders. In others, a significant concern is preventing denial of service attacks that involve disabling location reference nodes, exploiting vulnerabilities associated with the MAC protocol, or physical-layer jamming. Another key concern, the one that we focus on primarily in this work, is that of preventing degradation of the accuracy of the location service itself. This involves detecting, correcting, and mitigating the deliberate insertion of false signals (spoofing) by attackers.

We propose a novel methodology for secure localization that is inspired by error control coding techniques that have been used for many decades to provide robustness to noisy channels in traditional digital communications [8]. In error control coding techniques, the message to be sent is first encoded into a higher-dimensional codeword. When the message passes through a communication channel, it may be distorted due to noise and interference effects. At the receiver end, errors introduced by the channel can be detected (to an extent that depends upon the code-rate, the ratio of message bits to the bits in the codeword) by exploiting the fact that there are a relatively small number of correct/feasible codewords in the codeword space. Erroneous messages can be corrected, again to an extent depending upon the code-rate, by mapping the distorted encoded message to the the nearest feasible codeword and then decoding back to the original message.

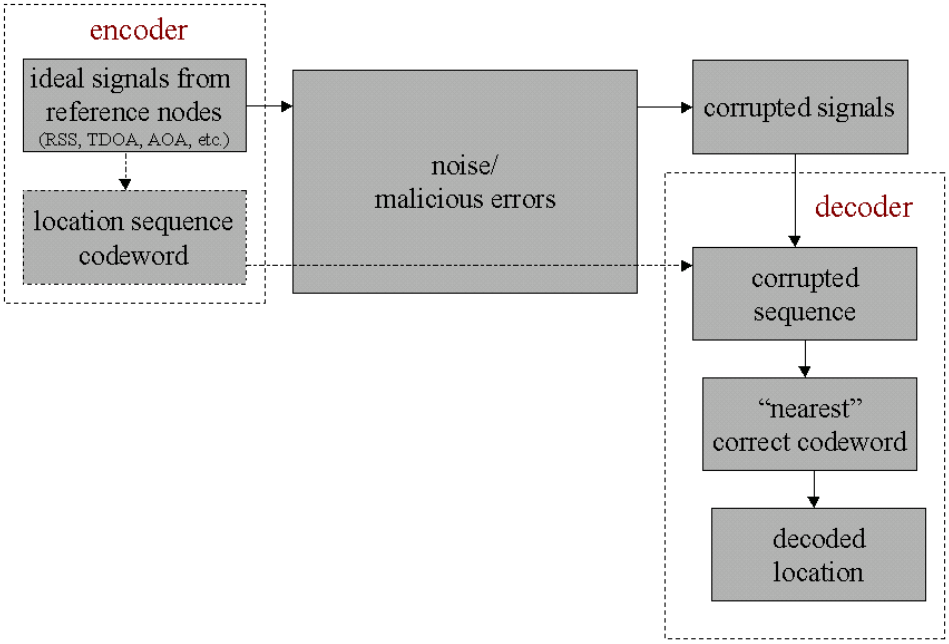

Fig. 1. Localization as analogous to error control coding

We present a sequence-based self-error-correcting localization mechanism, that is depicted as being analogous to error control coding in Figure 1. In this mechanism, the codeword that is used for localization is formed as a sequence by ranking the distances from the mobile node (with unknown location) to the various near-by reference nodes. For a given deployment of reference nodes, different location regions in the deployment area have distinct and unique

codewords that can be tabulated in advance. Given such a codeword, the location of the node is then determined by a reverse table-lookup.

A key to the security of this technique is that the the density of feasible codewords (i.e. the codewords that correspond to an actual location region in the deployment area) decreases steeply with the number of location reference nodes. In particular, if N represents the number of reference nodes in a two-dimensional area, only $O(N^4)$ of the $O(N!)$ possible sequences are feasible codewords. This low density of feasible location codewords provides the ability to detect, correct and mitigate errors in the encoded location sequence that may be deliberately introduced by malicious attackers by the spoofing of ranging signals.

We present a brief numerical evaluation of the robustness of the proposed technique to signal spoofing by a powerful attacker. While the results from this technique are indeed encouraging, it is possible that the particular mechanism we propose may be just a tip of the iceberg. Extensions of this technique may reveal a large space of related secure location coding techniques that provide even greater immunity to malicious attacks.

2 Related Work

Localization of nodes in the context of densely deployed wireless networks in benign settings has been the object of intense study in the past decade. Several articles have provided a thorough survey of the subject (e.g., see [2], [3, Ch. 3]). However, it has been only even more recently that researchers have developed localization techniques suitable for settings involving malicious attackers. We briefly survey some of this work on secure localization:

In [7], the authors propose a secure localization technique that offers protection against spoofing of reference node locations by attackers. In this technique, localization is performed using a set of public base stations as reference nodes, whose location claims are verified by another set of covert base stations. The probability of the attacker's success is shown to grow linearly with ranging error, inversely to the square-root of the area of the localization space, and inversely to the number of covert base stations. In [1], another work that studies secure localization against spoofing of locations of reference nodes, two techniques based on majority-vote are proposed. In the first technique, the spoofed location is assumed to appear as a "outlier" in the set of locations of all reference nodes. The outlier is detected and discarded by verifying the unknown node location estimate with different subsets of reference nodes. In the second technique, reference nodes vote on the unknown node location likelihood at each grid point in the localization space and the grid point with the highest number of votes is chosen as the location estimate.

A statistical method of secure localization is proposed in [9]. In this method, the authors suggest that using the median of localization data is inherently more secure than using its average. Based on this, they propose

that triangulation based localization techniques such as least sum of squares estimator should instead use least median of squares. They also argue that RF finger-printing based localization techniques should use minimum median distance instead of minimum average distance as the metric to determine the nearest neighbor.

In a work that addresses other attacker models, the authors in [4] propose a secure localization technique that is robust to *Wormhole* and *Sybil* attacks. In this technique, secure localization is provided using a combination of encryption keys, power control, and directional antennas that can transmit in different directions at the same time. In [5], the authors propose a combination of distance bounding, exchange of challenge-response type of messages, and authentication keys to ensure security of location determination. The authors of [6] address the problem of location verification in which the location of a wireless node is verified for its authenticity. This is done by using the inherent constraints in round trip propagation times of RF and ultra sound signals.

The technique that we describe here is quite different from prior work on secure localization in its use of ordered sequences as a location code. It is most similar to the Ecolocation technique that we previously proposed for localization using pure RF signals [10]. A key difference from that work is that we focus here on errors introduced by malicious attackers, rather than RF channel errors due to multi-path fading and other environmental factors.

3 Secure Sequence-Based Localization

Consider a two-dimensional location area in which N reference nodes with known locations are deployed. The goal is to enable a mobile node in the environment with an unknown location to locate itself. The mechanism can be implemented in either a node-centric or an infrastructure-centric mode. In the former implementation, the node obtains distance estimates to each reference node and computes its location according to the mechanism described below. In the latter, the reference nodes each obtain a distance estimate to the unknown node and transmit this information to a centralized node where the computation is performed.

The exact mechanism for obtaining the distance estimate can vary in either case. For instance, it could be based on time difference of arrival between a radio and acoustic signal, or based on an estimate using pure radio signal strength, or possibly even via phase interferometry). For a detailed discussion of how sequence-based localization provides robustness to errors in the distance estimation in the context of pure-RF localization, we refer the reader to the evaluation of the Ecolocation technique presented in [10]. In the discussion below, we shall assume that the ranging estimates are sufficiently accurate that they affect the accuracy of localization minimally. Thus our focus with regard to errors will be solely on those introduced by a malicious attacker.

3.1 Sequence Encoding of Locations

We now give a description of how the ranging signals are interpreted for localization in the deployment region. First, the distance estimates between the unknown node and the reference nodes is converted to a sequence that represents the relative distance ranking of each reference node (from nearest to farthest).

A correct sequence represents a valid location region in the deployment area where the distances to the different reference nodes are ranked accordingly. The two-dimensional deployment area can be partitioned into all valid location regions by drawing all lines that are perpendicular bisectors of lines joining all pairs of reference nodes. The two sides of each such line represent the distance inequalities with respect to a particular pair of reference nodes. The faces formed by the intersection of these lines represent all feasible regions, each forming unique combinations of the inequalities that result in the corresponding sequence.

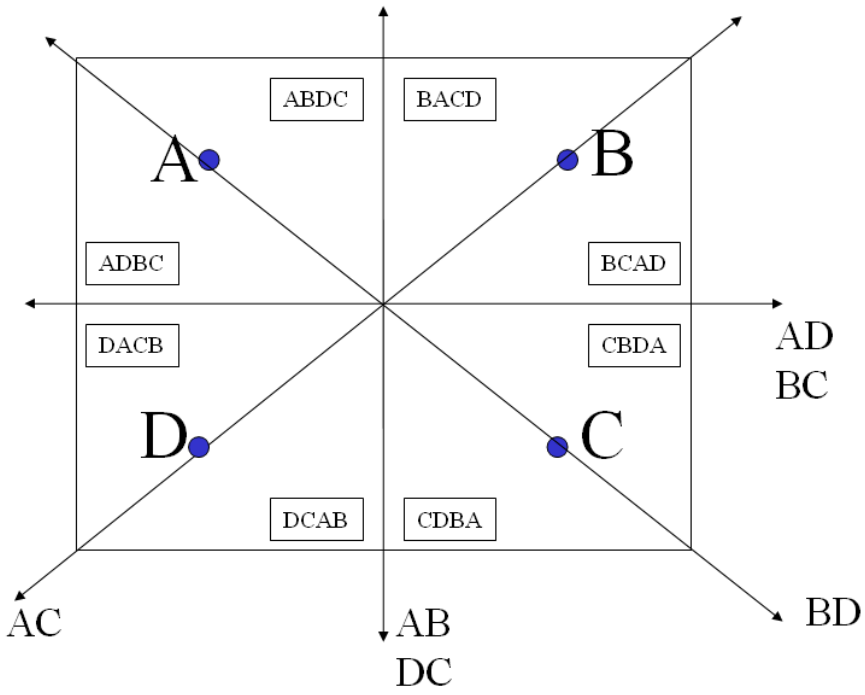

Fig. 2. Illustration of location regions and sequences for four nodes in a square region

For example, consider figure 2. It shows a square region in which there are four reference nodes labeled A, B, C, and D. Because the perpendicular bisectors for the node pairs AB, DC and that if node pairs AD, BC overlap respectively, there are four bisecting lines in all, dividing the region into eight location regions. The unique sequences corresponding to each of the eight regions area also shown. For instance the region on the bottom of the top-left quadrant corresponds to the sequence ADBC because all points in this region are closest to A, closer to D than B and C, and finally closer to B than C.

This mapping between location regions and location sequences can be encoded in a table *a priori*, during or before network initialization. In the absence of any errors, the unknown node can be located by performing a reverse look-up on this table to obtain the location given the sequence. Localization using a table look-up is quite feasible for intermediate-sized networks (we show in the following section that the size of the table is polynomial); the speed can be further improved using multi-resolution and gradient-based grid search techniques.

3.2 Density of Feasible Sequences

We now investigate a bound on the number of feasible sequences for a given deployment. Given that there are $N!$ possible sequences in all, how many of these actually correspond to feasible locations? The following proposition addresses this question, which is of core relevance not only to the time and space complexity of the algorithm, but also its ability to detect and correct errors.

Proposition 1. *In a two-dimensional area with N reference nodes, the number of feasible location sequences (i.e. sequence codewords that correspond to location regions in the 2D area) is $O(N^4)$.*

Proof: Each feasible location sequence represents a combination of pairwise inequality relationships with respect to the distance to two reference nodes. Each corresponding location region results from the intersection of these inequalities, each of which is represented by a line in the 2D plane that is a perpendicular bisector of the line joining each pair of reference nodes. It suffices to show that the number of faces formed by these intersecting lines is $O(N^4)$. We can do this in two steps.

- First, note that the number of bisecting lines is $O(N^2)$, since there is at most one line for each pair of reference nodes and there are $C_2^N = \frac{n(n-1)}{2}$ pairs of nodes.
- Next, we can prove that the arrangement of k lines in the plane can result in at most $O(k^2)$ faces. This can be proved by induction by showing that each line added to an existing arrangement of $i-1$ lines adds at most i additional faces, which bounds the total number of faces by $k(k+1)/2+1$.

Combining these together, we get that the total number of faces with N reference nodes must be $O(N^4)$.

$\diamond$

Note that we have assumed here a strict ordering of references, ignoring any lines or intersection points in the region where two or more references are equidistant, otherwise there could be potentially N^N total sequences (not $N!$), though the number of feasible sequences remains $O(N^4)$ even in that case.

3.3 Decoding Falsified Sequences

When there are no errors in the observed sequence, the localization algorithm is simple as it requires only a table look-up; it also provides excellent performance in terms of accuracy. Since there are $O(N^4)$ regions in a given deployment area A, the average area of location regions decreases roughly as $O(\frac{A}{N^4})$, providing a rapid gain in location accuracy as the number of reference nodes increases.

However, in the face of an attacker or even errors due to the environment, the original sequences may not be received correctly. The error detection and correction capabilities of the sequence-based localization mechanism are called for. These capabilities rely on the proposition we described in the last section: the ratio of feasible sequences (that correspond to valid locations in the deployment area) and the ratio of all possible sequences is very small for even moderate values of N, the number of reference nodes.

Most changes to the reference nodes' signals will not result in a sequence with a feasible location — this provides the basis for error detection. If the ranging mechanism used is inherently highly accurate (such as TDoA techniques in some settings), then a sequence that does not correspond to any of the feasible location region provides an indicator that a malicious attack may be in progress. Another way to view this is that the location encoding provides a consistency check because they represent fundamental geographic constraints; when a received sequence fails this consistency check, one can infer that errors have been introduced into the system.

Given that there is an error introduced into the sequence, a location can still be obtained from the sequence. This is done by mapping the erroneous sequence to the "nearest" feasible codeword sequence corresponding to a valid location region. The notion of "nearest" can be defined using different metrics suitable for sequences. A simple choice is to count the number of pairwise inequality constraints (pertaining to reference node distances) that are violated in the given erroneous sequence with respect to each feasible region, and pick the region that minimizes the number of unsatisfied constraints. Consider again the example deployment in figure 2. If the sequence ABCD is received, it can be detected immediately that an error has been introduced into the

sequence since this sequence does not correspond directly to any of the location regions. The nearest feasible codeword to this sequence is ABDC, which corresponds to the region on the top of the top-left quadrant. This is because there is a difference of only one constraint violation between these two sequences (the inversion of the 'D is closer than C' constraint). If this mapping is assumed to be correct based on limitations on the attacker's capability, this may suggest that the attacker is either manipulating the signals of C to make it appear closer to the mobile node, or the signals of reference node D to make it appear farther than it really is to the mobile node.

While there is no guarantee that this approach will necessarily result in a correct solution in all cases, it can mitigate significantly the impact of the errors in the location sequence. The likelihood of correction and the obtained accuracy both increase with the number of reference nodes.

4 Evaluation

We briefly evaluate the proposed sequence-based localization mechanism though a set of simulations involving malicious attackers.

Evaluating the performance of a localization mechanism with respect to a malicious attacker requires some care. In particular, we need to be explicit about the capabilities of the attacker and strike a balance between providing the attacker with sufficient power and imposing some costs for the attack. Our approach is to impose such a cost by limiting the number of reference nodes that the attacker can manipulate to k (we provide numerical simulations for $k = 1$ and $k = 2$ here, but these can be extended to other values of k). At the same time, we provide an advantage to the attacker by allowing the attacker to know the 'true' location of the mobile node and switch accordingly at each time to the set of k reference nodes that hurts the localization technique the most in terms of accuracy. This allows the attacker to impose a worst-case penalty on the performance of our localization mechanism at all locations in the deployment area. In this sense, the provided results represent an upper bound on the damage that could be suffered at the hands of an intelligent attacker. In particular, the location accuracy in real attack scenarios may be substantially better than what is shown in these figures. This is because realistically, in some settings, the attacker may not able to observe the unknown node, and therefore may not be able to change which reference node is being spoofed or manipulated on the fly as the mobile node moves through the network.

We measure location error as the absolute distance between the true and estimated location as a percentage of D_a, the average distance between pairs of reference nodes; this provides normalization with respect to the density of reference node deployment. In case of a single strong attacker, whose impact is evaluated in Figure 3, the attacker inflicts the most damage on the localization scheme by introducing $N - 1$ constraint violations; this is done by moving the

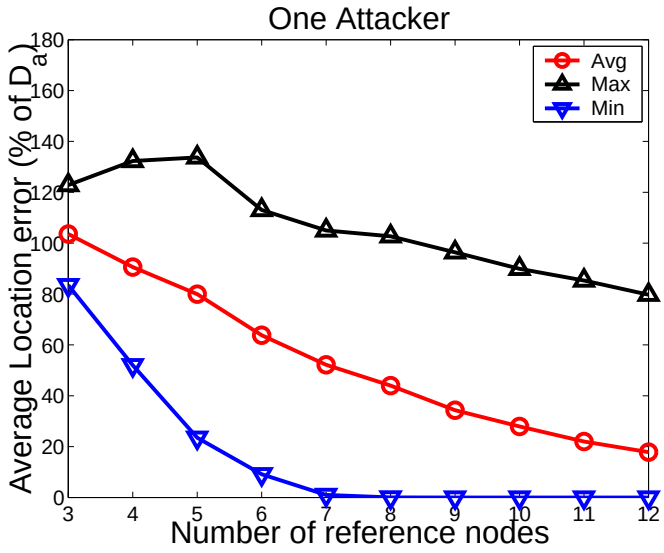

Fig. 3. Impact of one powerful attacker on secure sequence-based localization in a square area

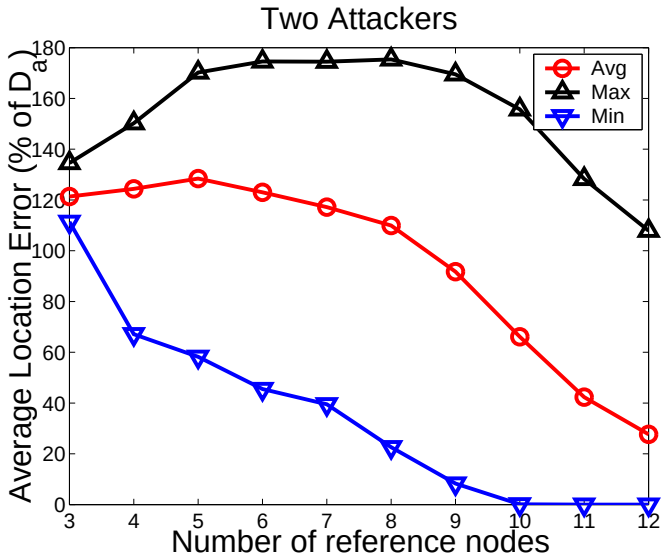

Fig. 4. Impact of two powerful collaborating attacker on secure sequence-based localization in a square area

first node on the sequence to the end of the sequence (by spoofing the nearest reference node to the mobile nodes' true location to make it seem the farthest). The three curves represent the normalized location error as a function of the number of reference nodes. The max, min and average shown are taken over all locations chosen uniformly from a square deployment area. The overall decreasing trends in all three curves is highly encouraging (although the max curve shows a small increase for less than five references). The steep falloff in the curve for min and the slower falloff for max suggests that there are some locations of the unknown node where a single attacker is quite powerless to affect the localization accuracy, and others where it has greater impact. On average, the proposed technique provides significant improvements (error less than 20% of inter-node spacing) when more than 12 reference nodes are present.

Figure 4 shows the corresponding results when two intelligent attackers collaborate to spoof the readings from two different reference nodes. In this case the worst damage to the location technique is inflicted by moving the first node in the sequence to the end, and bringing the last node in the sequence to the front (by adjusting the distance estimates for these two reference nodes accordingly). We observe again that there are decreasing trends in max, min and average location error once there are sufficiently many reference nodes (though there is an initial increase for a small number of references). We also see that there exist locations where two attackers have negligible impact on localization error when there are at least 10 references.

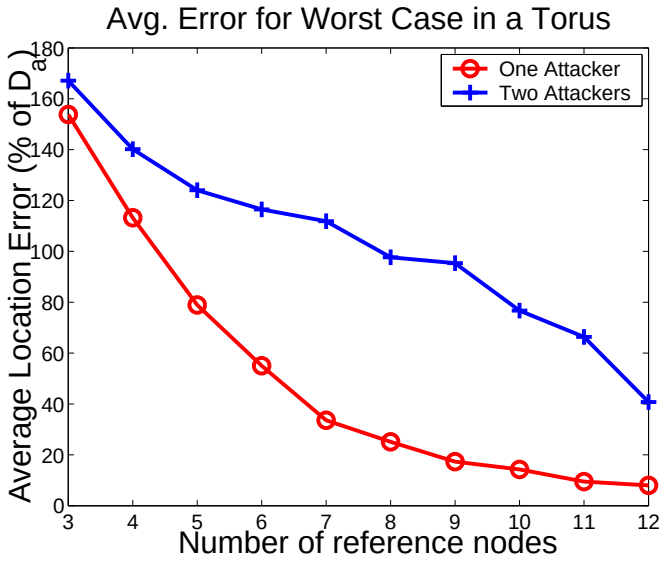

Fig. 5. Impact of one and two attackers on secure sequence-based localization in a torus (no edge effect)

When relatively small-to-medium numbers of reference nodes are considered, the location regions obtained tend to exhibit some skewed boundary effects (the regions near the boundary are larger than those in the center). Eliminating these edge effects gives an idea of the performance in the context of a seamless deployment of a large number of reference nodes (only a subset of which are within range at any location of the unknown node). We therefore examine the performance of the normalized location error (averaged over all locations) on a torus, for both single and two-reference-node attacks, in Figure 5. We observe that for the single attacker case, while the initial localization error for a torus is higher compared to the grid, the falloff is faster so that by the time there are 12 reference nodes, the localization error is lower (less than 10% of the baseline inter-reference spacing). For two attackers, the average localization error also shows a steady (almost linear) decline for the torus, though it appears slightly worse than in the case of the grid.

These simple experiments suggest on the whole that localization using sequences provides robust protection against malicious attacks in dense deployments.

5 Conclusions

We have presented a novel sequence-based secure localization mechanism. The key to its performance (which improves rapidly with increasing density of reference nodes) is that the number of feasible sequences is considerably smaller than the set of all sequences that can be generated, allowing for robust detection and correction of errors in the sequence. In future work it would be of interest to evaluate this technique for attacks involving an even greater number of compromised nodes, and compare the security provided by this technique with other state-of-the-art approaches. Still, this technique is perhaps just the tip of an iceberg; much remains to be learned about such self-error-correcting localization techniques, particularly with respect to malicious attacks.

6 Acknowledgements

The research in this work has been supported in part by National Science Foundation through grants NeTS-NOSS CNS-0435505, CAREER CNS-0347621, ITR CNS-0325875, and CCF-0430061, and through a gift from Bosch Research.

References

1. Donggang Liu, Peng Ning, and Wenliang Du. Attack-resistant location estimation in sensor networks. In *Proceedings of the Fourth International Symposium on*

Information Processing in Sensor Networks, IPSN, pages 99–106, Los Angeles, CA, USA, 2005.
2. Jeffrey Hightower and Gaetano Borriello. Location systems for ubiquitous computing. *Computer*, 34(8):57–66, August 2001.
3. Bhaskar Krishnamachari *Networking Wireless Sensors* Cambridge University Press, 2005.
4. Loukas Lazos and Radha Poovendran. SeRLoc: secure range-independent localization for wireless sensor networks. In *WiSe '04: Proceedings of the 2004 ACM workshop on Wireless security*, pages 21–30, New York, NY, USA, 2004. ACM Press.
5. Loukas Lazos and Radha Poovendran and Srdjan Capkun. ROPE: Robust Position Estimation in Wireless Sensor Networks. In *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks, IPSN*, Los Angeles, CA, USA, 2005.
6. Naveen Sastry, Umesh Shankar and David Wagner. Secure verification of location claims. In *WiSe '03: Proceedings of the 2003 ACM workshop on Wireless security*, pages 1–10, New York, NY, USA, 2003. ACM Press.
7. Srdjan Capkun, Mario Cagalj and Mani Srivastava. Secure Localization With Hidden and Mobile Base Stations. In *IEEE INFOCOM*, Barcelona, Spain, 2006.
8. Stephen Wicker, *Error Control Systems for Digital Communication and Storage*, Prentice-Hall, 1995.
9. Zang Li, Wade Trappe, Yanyong Zhang and Badri Nath. Robust Statistical Methods for Securing Wireless Localization in Sensor Networks. In *Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks, IPSN*, Los Angeles, CA, USA, 2005.
10. Kiran Yedavalli, Bhaskar Krishnamachari, Sharmila Ravula and Bhaskar Srinivasan. Ecolocation: A Sequence Based Technique for RF-only Localization in Wireless Sensor Networks In *Proceedings of The Fourth International Conference on Information Processing in Sensor Networks, IPSN*, Los Angeles, CA, April 2005.